LINKSCAPE · ETHICS

Responsible AI Policy

The rules we hold ourselves to when we build and release AI systems.

DELIVERABLE	E-01
STATUS	Adopted

Executive Summary

This document establishes the foundational principles and operational guidelines that govern artificial intelligence (AI) development, deployment, and oversight at LinkScape. As a youth-led nonprofit dedicated to responsible innovation in AI, LinkScape is committed to ensuring that all AI initiatives align with our core values of responsible innovation, transparency, accessibility, impact, and safety.

This policy serves as a comprehensive framework for: (1) establishing clear AI development principles; (2) ensuring fairness and bias mitigation throughout the AI lifecycle; (3) maintaining transparency in AI systems and decision-making; (4) implementing robust human oversight mechanisms; (5) defining accountability structures; and (6) establishing comprehensive AI safety guidelines.

All team members, partners, and collaborators working with LinkScape are expected to adhere to these principles and guidelines.

1 AI Development Principles

LinkScape's AI development is grounded in a set of core principles that guide decision-making at every stage of the AI lifecycle, from conception through deployment and ongoing maintenance.

1.1 Responsible Innovation

We believe that innovation must be pursued with intentionality and care.

- All new AI systems must undergo impact assessment before deployment
- Potential risks and benefits must be documented
- Development must balance innovation with risk mitigation
- We commit to iterative improvement based on real-world feedback

1.2 Transparency by Design

Transparency is fundamental to building trust with users, stakeholders, and the broader community.

- We will clearly communicate when AI is being used in decision-making processes

- Documentation of AI systems will be made available to stakeholders
- Limitations and uncertainty in AI systems will be clearly disclosed
- We will maintain up-to-date records of all AI systems in use

1.3 Accessibility and Inclusivity

AI technology should benefit all members of our community, regardless of background or technical expertise.

- AI systems will be designed with accessibility as a core consideration
- Diverse perspectives will be incorporated in system design and testing
- We will actively seek feedback from underrepresented communities
- Training and educational resources will be made widely available

1.4 Measurable Impact

We are committed to creating AI systems that generate tangible, positive value.

- All major AI initiatives must have clearly defined success metrics
- Impact will be measured against our organizational mission
- Progress will be tracked and reported regularly
- Systems will be refined or discontinued if they fail to achieve intended impact

2 Fairness and Bias Guidelines

Bias in AI systems can perpetuate discrimination and harm vulnerable populations. LinkScape is committed to identifying, measuring, and mitigating bias throughout the entire AI development lifecycle.

2.1 Bias Assessment and Mitigation

Every AI system must undergo rigorous testing for bias:

- Data Audits: All training datasets must be audited for representativeness and potential biases
- Performance Testing: Performance metrics must be evaluated across different demographic groups
- Adversarial Testing: Systems will be tested against adversarial examples to identify failure modes
- Documentation: All bias testing results will be documented and reviewed by governance teams

2.2 Diversity in Development Teams

Diverse teams create more robust and equitable AI systems:

- Development teams will intentionally include diverse perspectives
- Code reviews will include consideration of fairness implications

- Cross-functional collaboration will include ethics and social impact considerations

2.3 Equitable Access and Outcomes

We will work to ensure that AI systems serve all users fairly:

- AI services will be made available to underserved communities
- We will monitor for and address disparities in system outcomes
- Feedback mechanisms will be in place to identify and remedy unfair outcomes

3 Transparency Requirements

Transparency enables accountability and builds public trust. LinkScape commits to maintaining clear, comprehensive documentation and communication about our AI systems and their impacts.

3.1 AI System Documentation

Each AI system must have complete documentation that includes:

- System Overview: Purpose, capabilities, and limitations
- Technical Specifications: Architecture, algorithms, and key parameters
- Data Documentation: Training data sources, characteristics, and collection methods
- Performance Metrics: Accuracy, fairness metrics, and performance across groups
- Risk Assessment: Identified risks and mitigation strategies
- Use Cases and Limitations: Intended uses and explicit out-of-scope applications

3.2 Public Communication

We will maintain open communication with the public:

- AI system disclosures will be made in plain language accessible to non-experts
- Users will be informed when they are interacting with AI systems
- We will publish regular reports on AI system performance and incidents
- Feedback channels will be maintained for public input and concerns

3.3 Data and Algorithm Transparency

We will maintain transparency regarding data and algorithms:

- Data sources and collection methods will be documented
- Algorithm descriptions will be made available where proprietary considerations permit
- Explainability methods will be employed for high-impact decisions
- Updates to systems will be communicated to stakeholders

4 Human Oversight Procedures

Human judgment and oversight are essential to responsible AI deployment. LinkScape maintains meaningful human involvement in AI decision-making, particularly for high-stakes applications.

4.1 Levels of Human Oversight

AI systems will be classified by impact level, with corresponding oversight requirements:

Impact Level	Description	Oversight Requirements
Low	Administrative/routine decisions	Logging; periodic review
Medium	Individual-level decisions with significant impact	Human review before deployment; continuous monitoring
High	Decisions affecting rights, welfare, or opportunities	Human-in-the-loop; approval required; regular audits

4.2 Decision-Making Processes

Human involvement in AI decisions will follow these principles:

- Explainability: Humans must understand the reasoning behind AI recommendations
- Contestability: Affected individuals must have meaningful recourse to challenge AI decisions
- Accountability: A human must be responsible for final decisions in high-impact contexts
- Appeal Mechanisms: Clear processes for appeals and human review of disputed decisions

4.3 Monitoring and Intervention

Continuous monitoring ensures systems remain responsible:

- Performance monitoring: Continuous tracking of accuracy and fairness metrics
- Anomaly detection: Systems to identify unusual patterns or failures
- User feedback: Mechanisms to capture and act on user concerns
- Intervention protocols: Clear procedures for immediately disabling systems when necessary

5 Accountability Framework

Clear lines of accountability ensure that AI development is responsible and that problems are addressed promptly. LinkScape establishes specific roles and responsibilities for AI governance.

5.1 Organizational Roles

Leadership and governance structure:

- Co-Founder & CFO (Liqian (Eric) Yan): Overall responsibility for AI governance and strategic alignment
- AI Ethics Lead: Responsible for policy development and compliance
- Development Teams: Responsible for implementing policies in daily development
- Ethics Review Board: Independent review of high-impact AI decisions

5.2 Incident Response and Remediation

When AI systems cause harm or fail, LinkScape commits to:

1. Rapid Response: Immediately assess the scope and impact of the incident
2. Transparency: Communicate openly with affected parties within 72 hours
3. Investigation: Conduct thorough root cause analysis
4. Remediation: Implement corrective measures to prevent recurrence
5. Accountability: Hold responsible parties accountable
6. Learning: Share findings and learnings across the organization

5.3 Regular Audits and Assessments

LinkScape will conduct regular comprehensive audits:

- Quarterly reviews of all active AI systems
- Annual comprehensive fairness and bias audits
- Third-party audits of high-impact systems
- Documentation and reporting of all findings

6 AI Safety Guidelines

AI safety encompasses the technical and operational measures necessary to ensure AI systems operate reliably, securely, and in alignment with their intended purposes. LinkScape prioritizes safety at every stage of AI development and deployment.

6.1 Development Practices

Safe AI development requires rigorous practices:

- Code Review: All AI code undergoes peer review with explicit safety focus
- Testing Protocols: Comprehensive testing including edge cases and adversarial examples
- Version Control: All models and code maintained with proper version control
- Documentation: Complete documentation of development processes and decisions
- Security Review: Assessment for security vulnerabilities before deployment

6.2 Data Security and Privacy

Protecting data is essential to responsible AI:

- Data Protection: Access restricted to authorized personnel; encryption of sensitive data
- Privacy by Design: Systems designed to minimize personal data collection
- Consent Management: Explicit user consent obtained for data use
- Data Retention: Clear policies for data deletion when no longer needed
- Audit Trail: All data access logged and auditable

6.3 Robustness and Reliability

AI systems must be robust and reliable:

- Performance Monitoring: Continuous tracking of system performance in production
- Degradation Detection: Rapid identification and alerting of performance degradation
- Failover Mechanisms: Automatic fallback to safe alternatives when systems fail
- Dependency Management: Clear understanding and management of system dependencies
- Update Procedures: Safe processes for updates and model refreshes

6.4 Adversarial Robustness

Systems will be tested and hardened against adversarial attacks:

- Adversarial Testing: Regular testing against known attack vectors
- Robustness Metrics: Tracking of system robustness across perturbations
- Input Validation: Careful validation and sanitization of all inputs
- Defense Mechanisms: Implementation of defense strategies against known attacks

Conclusion

This Responsible AI Policy reflects LinkScape's commitment to developing and deploying AI systems that are fair, transparent, safe, and beneficial to society. As a youth-led organization at the forefront of responsible AI innovation, we recognize our responsibility to set high standards for the field.

Our core values of responsible innovation, transparency, accessibility, impact, and safety guide every decision we make. By adhering to these principles and guidelines, LinkScape aims to demonstrate that AI can be developed in ways that respect human rights, promote fairness, and generate meaningful positive impact.

We commit to regularly reviewing and updating this policy as the field of AI evolves and as we learn from our experiences. We welcome feedback from our community, partners, and stakeholders, and we remain committed to continuous improvement in all aspects of our AI governance.

LinkScape Team

Last Updated: 2/13/2026

Appendix: Definitions and Key Terms

This appendix provides definitions of key terms used throughout this policy.

Artificial Intelligence (AI)

Software systems designed to perform tasks that typically require human intelligence, such as learning from data, recognizing patterns, understanding language, or making decisions.

Bias

Systematic errors in AI systems that result in systematically different outcomes for different groups. Bias can arise from training data, algorithm design, or deployment contexts.

Fairness

The property of treating individuals and groups equitably, without systematic disadvantage. Fairness in AI means systems should not discriminate based on protected characteristics.

Explainability

The ability to understand and articulate the reasoning behind an AI system's decisions. Explainable systems allow users to understand why a particular decision was made.

Human-in-the-Loop

An AI system design where humans remain involved in the decision-making process, providing oversight and validation of system recommendations.

Transparency

The disclosure of relevant information about AI systems, including their purposes, capabilities, limitations, and impacts. Transparency enables accountability and informed decision-making.